

# 知的情報検索の応用

## Applications of Intelligent Informaion Retrieval

篠原 靖志\*  
Yasushi Shinohara

堤 富士雄\*  
Fujio Tutumi

\* (財) 電力中央研究所  
Central Research Institute of Electric Power Industry.

1998 年 5 月 25 日 受理

**Keywords:** information retrieval, information filtering, user interface.

### 1. ま え が き

1990 年代のパソコンの浸透とネットワーク化, 特に, WWW の登場は, 膨大な量の多様な電子情報の氾濫と, 情報検索の利用者数の急速な増加を引き起こしている. 現在, 索引可能な Web ページ数は, 32000 万ページ [Lawrence 98] に達すると言われている. これに伴い, インターネットやイントラネット上で情報検索を行う利用者数も増加し, また, 多様になってきている. 現代の情報検索の欲求は, この情報の爆発が前提であり, 知的情報検索の課題は, この多様な内容を持つ膨大な文書の中から, 多様な利用者が必要とする情報を, いかに容易かつ的確に提供するかである.

本稿では, 利用者の多様な要求に対して, 容易かつ的確に情報を提供するための検索技術の応用の現状を, 利用者とのインタラクションという視点から概観したのち, インタラクションの仕方の異なる 2 つの実用的システムを紹介する. 最後に, 知的情報検索におけるインタラクションの課題について述べる.

現在, 音声・画像などテキスト以外のマルチメディア情報も増加しつつあり, これらの情報の整理・検索の重要性は今後増してくると考えられるが, 本稿では, テキストを中心とする文書検索を対象とする.

なお, 知的情報検索の概要と研究開発の動向については [住田 96] に, また, 知的情報統合の観点からいくつかの実験システムの紹介が [武田 96] に, 解説されている.

### 2. 知的情報検索と利用者とのインタラクション

利用者が与えた検索条件のみから, 利用者の意図を推定し, いかに質の高い検索結果を返すかは, 知的情報検索の 1 つの重要で基本的な課題である.

しかしながら, 利用者は自身の要求を検索条件として十分に表現して, 検索システムに提示できるとは限らない. このような状況としては, 利用者の関心がまだ漠然としているために, 自分の要求を明確に述べる事ができない状況や, 調べたいものは比較的明確であるが, データベース中の用語の特性などが分からないために, 不適切な検索条件を使用してしまう場合などがある.

従って, 利用者が自力で記述した検索条件のみから, 必要とされる検索結果を返すには, 自ずから限界がある.

どのような情報を必要としているかを最も良く知り, 判断できるのは利用者であるが, その意図は必ずしも明示的ではない. このため, 知的情報検索の実現においては, 検索システムから返す情報とそれに対する利用者からの応答という双方向のインタラクションによって, 質の高い検索結果を得ることを目的とした支援技術が重要である. そして, 現在, このような双方向のインタラクションを積極的に活用する技術が, 実際の検索サービスにおいて活用されるようになってきている.

本章では, 商用または, それに近いレベルで実用化されている文書検索システムの基礎技術を利用者とのインタラクションの面から概観する.

### 2・1 検索条件に対する検索結果の質の向上

利用者から与えられた条件のみに基づいて、質の高い検索結果を得る技術は、利用者とのインタラクションを円滑にするための基礎技術である。

例えば、検索条件と関連度が高い順に検索結果を整理すると検索精度は向上する。このために、検索条件と文書との関連度を自動的に評価する技術としては、データベースから得た統計情報を用いて、重みとして *tf* や *idf* を用いたベクトル空間法 [Salton 89] が基本的技術となっている。この技術は、1970年代から研究されてきたが、現在も多くの改良が行われ、実用システムに組み込まれている。例えば、文書とキーワードの関連を考慮した改良である LSI(Latent Semantics Indexing)[Letsche 97] は代表的商用サービスの 1 つ Excite(<http://www.excite.com>) で使用されている。

適用分野を限定し、その領域モデルを構築することで、検索の精度を上げている実用システムもいくつか登場している。例えば SavySearch[Wove 97] は、複数の検索エンジンの使い分けについて、検索結果に基づいて学習したヒューリスティクスを適用して、検索精度を向上させている。

形態素解析など自然言語処理技術も基礎的技術である。単語索引を使用する検索システムでは、日本語の文の適切な単語分けが検索能力を大きく左右するが、会話体などにも適用可能な頑健な形態素解析システムを開発することで、検索精度を向上させている日本語用検索システムもある [井上 97]。

### 2・2 ナビゲーション情報の提示

利用者が次にどのような行動をとれば良いか、その選択の判断材料となるナビゲーション情報を提示することは、利用者の的確な応答を助ける。

主なナビゲーション情報としては、文書の抄録、関連キーワード、文書クラスタ、キーワードの分布、文書間の関係などが利用される。

特に関連キーワードの提示機能は、WWW 上の一般向けサービスでも普及しつつある。主に文書間のキーワードの共起性に基づいて、入力キーワードに関連したキーワードや絞り込みに有効なキーワードを提示する。LSIでのクラスタリング(Excite)や相関ルール[西村 96]などの手法により実用レベルに達している。

各種ディレクトリサービスは、文書のクラスタリングにより検索をナビゲートする代表であるが、Scatter/Gather[Hearst 95, Hearst 96]の考え方に基づいて、動的クラスタリングと関連キーワードの提示によ

って、よりインタラクティブ性を高めたサービスも登場してきている(例:クラスタリング機能付き TITAN <http://sting.navi.ntt.co.jp/titan/titan-clx.html>)。

ナビゲーションにおいては、提示方法も重要である。特に、2次元、3次元のグラフィックスを利用して、人間の高い視覚的情報処理能力を活用する実用システムが登場しつつある。例えば、商用サービス AltaVista(<http://altavista.digital.com>) の refine 機能では関連キーワードの依存関係がグラフ表示される。

より進んだ検索結果の視覚的表示方法に関しては Information Theatre[Rao 95]で多様な検討がなされている。例えば検索キーが検索結果の文書中でどういった分布をしているかを濃度の違うタイルの集まりで表示した TileBars、文書間の関連を屏風状に立体表示する Butterfly、情報の階層を接続された回転するドラム状に表示した Cone Tree などがある。例えば、Cat-A-Cone [Hearst 97] では Cone Tree を利用した検索結果の表示が検討されている。

### 2・3 利用者応答からの検索要求の推定

利用者が自ら要求を詳しく表現できない場合、および、検索条件入力の手間を省く場合などに、利用者の検索行動から利用者の要求を推定する技術は有効である。

代表的技術としては、関連性フィードバック(relevance feedback)がある [Rocchio 71]。これは、検索結果が検索目的に合っているか否かの利用者の評価(適合度評価)に基づいて、検索条件(キーワードとその重み)を自動的に変更するものである。1つの検索セッションで使用される場合もあれば(例:InfoBee[井上 97])、個人が関心を持つ情報を長期にわたってフィルタリングする場合にも使用されている。後者の例には、個人向けに関連新聞記事をフィルタリングして提供するサービスがある。

### 2・4 連続的なインタラクション

近年の計算機の処理高速化により検索に要する時間は短くなり、提示する情報の種別および対象とするデータベースの規模によっては完全な連続応答が可能となった。

利用者とのインタラクションが高速になってくると、利用者は時々刻々変化して提示される情報を見ながら、その変化を理解することで検索を効率的に行うことが可能になる。

例えば、大体の日付しか分からない情報を探す場合に、時間を示すメータの連続的な変化に応じて表示情報が連続的に変化するインタフェースを用意することで、利用者は大体の日時の見当から、時間のメータを

動かして情報を見ている内に、該当情報にたどり着くことができる。

こういった迅速なインタラクションを活用する手法としては、動的検索 (DynamicQuery)[Shneiderman 94] や、なめらかインタフェース [水口 95] が提案されており、実用的な検索アプリケーション開発に活かされている [Greene 98]。2つの手法の特徴は、連続的な検索条件の変更に対して連続的 (理想としてはリアルタイム) に直感的な反応を返すというものであり、人間の時間軸方向の認知能力が高いことを利用している。

このような迅速なインタラクションを実現する上では、例えば、動的検索システム Film Finder[Ahlberg 94] などのように前述の視覚的情報提示などを活用して、人間の認識力を高めたり、Timeline[Allen 98] におけるスライドバーの使用など、条件を容易に変化させることができる空間的情報入力などを活用する必要がある。

### 3. パーソナル電子新聞 ANATAGONOMY

本章では、関連性フィードバックによって利用者の検索要求を獲得するシステムとして、パーソナル電子新聞 ANATAGONOMY[坂上 96, Sakagami 97] を紹介する。

ANATAGONOMY は、利用者の興味を表すユーザプロフィールを関連性フィードバックによって学習することで、電子新聞の記事の中から、利用者の関心度 (スコア) が高いと予想される記事を選んで自動的に構成した紙面を提供する WWW 上の情報サービスである。

ユーザプロフィールは、キーワードとその重みのセットとして表現される。各記事に対する予想スコアは、ベクトル空間法による。即ち、ユーザプロフィールに対応するキーワードの加重ベクトルと、文書ベクトルの類似度によって決まる。

本システムのサーバは、ユーザプロフィールに基づいて、各記事に対する予想スコアを計算するスコア計算エンジンと、利用者固有のユーザプロフィールを学習する学習エンジンとからなる。

本システムのクライアントは、利用者の WWW ブラウザにダウンロードされた Java Applet であり、サーバから記事データを読み取り、記事に対する予想スコアを用いて自動的にレイアウトする。これによって、利用者の興味を反映した紙面が構成される。紙面構成には、新聞風表示モードとテレビ風表示モード、番付風表示モードなどがある (<http://www.labs.nec.co.jp/>

freesoft/ANATAGONOMY/).

図 1 に ANATAGONOMY の画面例を示す。各記事についてのスコアバー (score bar) が各記事のスコアを示す。その初期値はスコア計算エンジンからの値によるが、利用者がスコアバーを操作することにより明示的に修正できる。

#### 3.1 利用者の興味の抽出方法

ユーザプロフィールの学習では、紙面に表示された各記事に対して与えられたスコアと予想スコアとの差をなくすようにキーワードの加重を修正している。学習アルゴリズムは [中村 96] による。

問題は、どのようにして利用者の記事の閲読操作から各記事のスコアを決定するかである。ANATAGONOMY では、ユーザプロフィールに基づく記事の予想スコアを以下の 2つの方法で修正する。

1つは明示フィードバックであり、利用者が、各記事についてのスコアバーを操作してスコアを修正する。

もう1つは暗黙フィードバックである。これはシステムが利用者の記事の閲読操作に基づいて、スコアを自動的に修正するものである。ANATAGONOMY では、閲読の際の、画面拡大操作とスクロール操作を行った記事はユーザの関心が高いと仮定しており、一定のボーナス点が加算される。逆に、上記両操作を行わなかった記事に対してはボーナス点が減算される。これによって、ユーザは毎日の購読操作の積み重ねを通して自分の嗜好をシステムに通知することになる。

#### 3.2 評価実験

本節では、「暗黙フィードバック」、「明示フィードバック」の効果についての評価実験について述べる [坂上 96]。

実験は、フィードバック方式が異なる設定をした、以下の 4 グループに対して行われた。

- グループ A : 明示フィードバックのみを用いる。全ての記事に対してスコアバーを用いて点数を付ける。
- グループ B : 10 点のボーナス点による暗黙フィードバックのみを行う。記事のスクロールや拡大操作に 10 点を加点して、学習エンジンに返す。同一記事への重複操作は数えない。スクロールと拡大両方を行った記事に対しては 20 点を加点する。どちらの操作も行わなかった記事は 10 点減じる。
- グループ C : 30 点のボーナス点による暗黙フィードバックのみを行う。
- グループ D : 10 点のボーナス点による暗黙フィードバックと、スコアバーを用いた明示フィードバ

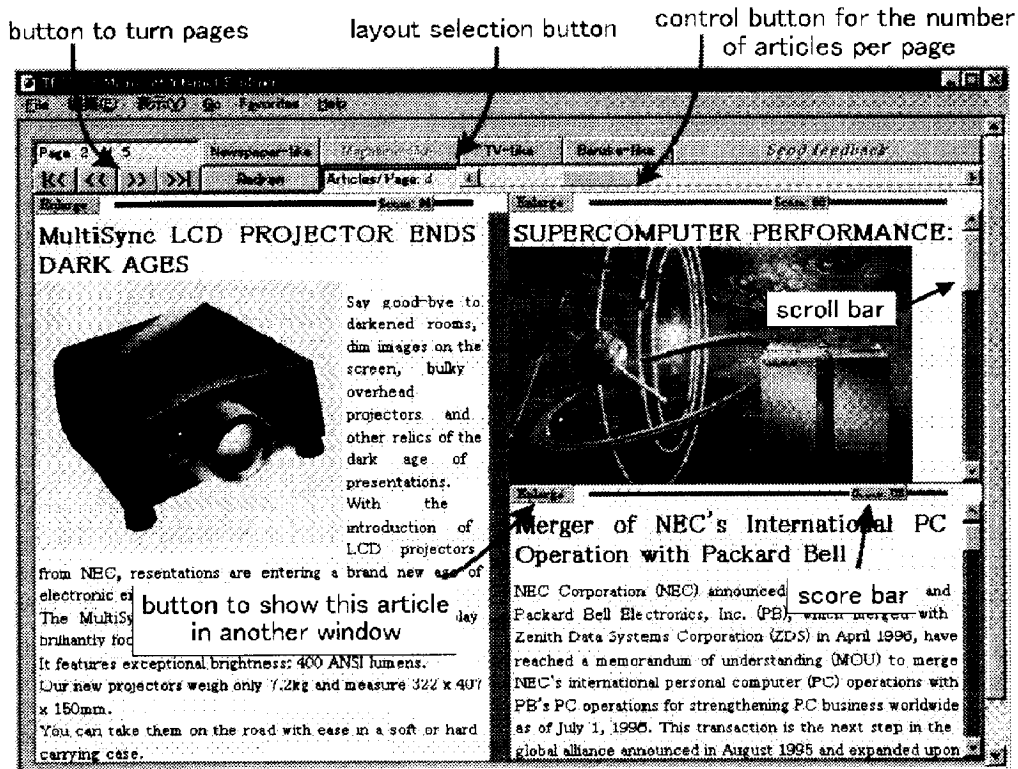


図1 ANATAGONOMYの検索画面例 [Sakagami 97]

クとの双方が可能である。

システムスコアに基づく順位とユーザの評価に基づく順位との差の平均の学習日数による変化を比較すると、グループA、グループD、グループB、グループCの順に、良く利用者の好みを学習していることが分かる。6日目になるとグループAとグループDの差はなくなっている。グループAで明示的にユーザがスコア操作を行った記事は平均18件であったのに対し、グループDでは平均6件であった。この結果、暗黙フィードバックと明示フィードバックを組み合わせることで、明示操作が少なくても、正確かつ迅速にパーソナライズできるとしている。

### 3.3 利 用 例

本システムは、現在、インターネット電子新聞サービス読売COLiNS(<http://pnews.cplaza.ne.jp/>)で使用されている。

### 3.4 ま と め

パーソナルなカスタマイズが可能な情報サービスは、PointCast(<http://www.pointcast.com>)以降、多数登場している。新聞のパーソナライズ化は、利用者の関心が長期間大きく変化しないという前提が成立しやす

く、関連性フィードバックなどの学習アルゴリズムを使用しやすい。

検索結果の適合度の判定をすべて利用者に明示的に行わせるのは、漠然とした関心を持つ利用者にとっては負担が大きい。本システムのように、利用者の購読操作からの暗黙の適合度評価と組み合わせる手法は、現実的な解決策となっている。

## 4. 文書検索システム「みつけるぞう」

本章では、利用者による迅速なインタラクションとナビゲーション情報の提示により、利用者の検索に対する制御性を高めることで、的確な検索を実現することをねらった「キーワードの2次元配置による文書検索システム」を紹介する。

### 4.1 概 要

本システムは、下記の3つの基本的機能を持つ。

- (1) ベクトル空間法に基づく検索
  - (2) つづりの類似に基づく曖昧検索
  - (3) 関連性フィードバックによるキーワード提示
- これらの機能の中の制御因子は、従来の検索システムでは、ほとんど自動的に決定されていた。

例えば、ベクトル空間法では一般に、検索結果は  $tf \cdot idf$  の和（または、その変形）によって整列される。

$tf$  (*term frequency*) は、キーワードの文書での出現頻度であり、 $idf$  (*inversedocument frequency*) は、キーワードのデータベースでの出現頻度の対数の逆数である。これらは、各々、キーワードの文書での重要度 ( $tf$ ) と、キーワードの検索における優先度 ( $idf$ ) を、自動的に推定したものと考えられる。

また、つづりの曖昧検索では、ニューラルネットワークや部分列インデクスを使用した各種システムがあるが、どの程度の類似性を許容するのかは、システムが自動的に決定している。

しかし、自動的に設定されたこれらの値が必ずしも利用者の意図と一致するとは限らない。本システムでは、これらの制御因子を、利用者が主体的に制御することができる。その制御にあたっては、迅速なインタラクションおよびナビゲーション情報の提示を工夫した設計がなされている。

#### 〔1〕 検索条件の空間的入力方式

本システムでは、キーワードを四角形の2次元平面上に配置することで、検索条件を表す(図2)。

2次元平面の各方向は、下記のような意味を持つ。

横軸は各キーワードの優先度を示す。これは、複数の検索キーワードがある場合の検索結果の整列の順序を決定し、優先度の高いキーワードを多く含む文書ほど上位に整列される。キーワードは、2次元平面の左に配置されるほど優先度が高く、右に配置されるほど低い。右端の排除の領域に置かれると、そのキーワードを含む文書は排除される (NOT の働き)。

利用者が最初にキーワードを入力した時点では、本システムはそのキーワードの優先度として、 $idf$  に相当する値を設定し、対応する位置にキーワードを配置する。このため、自動配置を操作しなければ、従来のベクトル空間法の検索結果と同等の結果が得られる。検索結果におけるキーワードの優先度の順番が不満足な場合、利用者は、キーワードをドラッグして、自由にキーワードの優先度を変えることができる。

一方、縦軸は、利用者が与えたキーワードのつづりの正確さを示す。キーワードを上端に配置すると、そのキーワードを含む文書を検索する。キーワードを下方に移動すると、そのキーワードと類似の文字列を含む文書が検索される。文字列の類似度は、含まれる  $n$ -gram (長さ  $n$  の文字列) の共通度によって定義する。1文字で類似用語を絞り込める度合いは字種により異なる。このような違いを吸収するため、漢字では1文字を、カタカナでは2文字列、アルファベットでは3文字

列を、 $n$ -gram として使用する。例えば、「ファジー」に対する「ファジイ」の類似度は、「ファ」「アジ」「ジイ」という3つの2文字列の内、「ファ」「アジ」の2つが共通に現れるので  $2/3 = 67\%$  となる。

利用者が最初にキーワードを入力した時点では、システムは、そのキーワードの正確さは100%と仮定して、入力平面の最上位に配置する。利用者は、自分の入力したキーワードのつづりが不正確であったり、つづりの似た別の書き方でデータベースに記述されている可能性がある場合は、キーワードをドラッグして下位に移動する。例えば、「ファジー」をつづりの正確さ60%の辺りに配置することで、「ファジイ」や「ファジイ推論」などのキーワードが類似キーワードと判定され、これらを含む文書が検索されることとなる。検索結果の順序は、横方向の配置から決まる優先度による。図2では、「気候変動」の曖昧さを高めることで、「気温変動」「気候変化」などが検索されている。

#### 〔2〕 キーワード候補の提示

本システムでは、2つの方法でキーワード候補の提示を行う。

1つは、キーワードを入力したり、配置した時点で、そのキーワードの類似用語と判定されたキーワードのリストが提示される(図2画面左下)。類似用語の判定は入力方式で述べた方法による。さらに、複数のキーワードが検索条件に配置されている場合、どのキーワードと共に、その類似用語が文書中で使用されているかという共起情報も提示される。これにより利用者は、類似用語の中から、現在の目的に合ったよりの確な検索キーワードを見つけることができる。

第2の方法は、利用者が適切だと考える文書(適正文書)を複数選択する方法である。本システムは、それらの文書に共通に現れる文字列を抽出して、これらをキーワード候補としている。これらを用語分析マップと呼ぶ2次元の平面に視覚的に提示することで、利用者がキーワードの適切さを把握しやすくしている。この場合の縦軸は、利用者が指定した文書での各キーワードの出現状況を、横軸は、各キーワードのデータベースでの出現頻度を示す。

#### 〔3〕 連続性の高いインタラクション

本システムは、2・3節で述べたような連続性の高いインタラクションを実現しようとしている。しかし、インターネット環境では、バンド幅の問題もあり、リアルタイムのフィードバックは困難である。本システムでは、利用者がキーワードの入力や配置変更が終了した度毎に、再検索を行い、即座に検索結果を更新することで、比較的低い計算量と通信量で連続性の高いイ

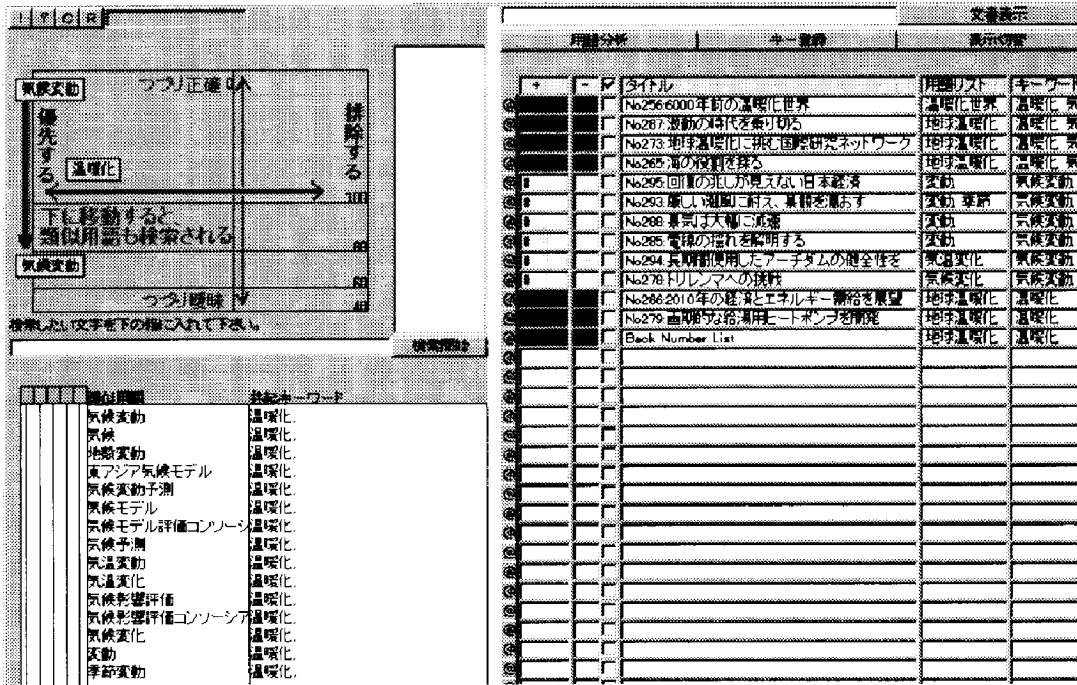


図2 キーワードの2次元配置による文書検索システムの画面例

インタラクションを実現している。

このため、例えば、キーワードを下方に移動していくと、類似用語が徐々に増えていく様子を見ながら、適切な所でキーワードの曖昧さを増すことを中止するといったことが行える。

〔4〕 その他

本システムでは形態素解析は行わず、漢字・カタカナ・アルファベットの連続した文字列を用語として取り出し、索引を作成している。平仮名列は索引されないが、平仮名のみによる用語をキーワードとする場合は少ないと仮定している。本方式では、辞書を必要とせず広く範囲の文書に対して索引作成を行える。

4.2 評価

本システムは、SearchSpace[提 98]と ExtractRequest[篠原 96, 篠原 98]という2つの実験システムを、その評価実験に基づいて、改良・統合し、Java アプレットとして実現したものである。ここでは、この評価実験について述べる。

被験者は、検索の非専門家である男女11名であり、実験対象文書は、主に、新聞朝刊1年分(記事5万本)である。評価対象システムの機能は下記のとおりである。

VisualWeightは、横軸方向の優先度操作のみによる検索システムである。SearchSpaceは、VisualWeightに曖昧検索機能を加えて、2次元配置による検索を行

システム	検索方式/機能	条件入力方式
VisualWeight	優先度検索	キーワードの1次元配置
SearchSpace	VisualWeight+曖昧検索	2次元配置
ExtractRequest	VisualWeight+用語マップ	1次元配置
WAIS-J	自動分ち書き+自動優先度設定	文章、単語列の入力
AND-OR	ブール検索	条件式入力

う。ExtractRequestは、VisualWeightに「用語マップ」の機能を付加したものである。AND-OR, WAIS-Jは、代表的な従来型検索方式である。

上位20位以内の正解率を比べると、従来型の2システムAND-OR, WAIS-Jが、30%程度であったのに対して、VisualWeight, SearchSpace, ExtractRequestの3システムは50%程度で、検索結果の上位に正解文書をより多く集めている。

1問当たりの検索回数(制限時間5分)はWAIS-J, AND-ORが各々3.9, 6.2であり、VisualWeight(9.6), SearchSpace(17.1), ExtractRequest(8.7)に比べて少ない。検索精度を考え合わせると、従来型2システムでは、利用者が検索結果に基づいて、検索条件を訂正するための情報や、容易な手段が乏しいことが考え

られる。その他、検索精度に差が出た理由としては以下が挙げられる。

まず、ブール検索による AND-OR では、正解を全く含まない検索結果を得る場合が多い(平均 6.2 回の検索中 3.3 回)。これは、「かつ」、「または」の使い分けを行えずに、不適切な検索条件を使用している場合が多々見られた。

一方、WAIS-J を使った場合は、検索結果に必ず無関係な文書が含まれた。複合語の自動分ち書きによる細分が原因である。また、これらを適切に制御して、ゴミを取り除く簡単な手段が利用者に与えられていない点による。

この実験から、連続的インタラクションなどの実現により、利用者が適切に制御因子を制御できることにより、検索の質が向上したと考えられる。一方で、1 問当たりの検索回数を考えると、漠然とした関心のあるテーマについて情報を集める利用者よりも、比較的明確な目的を持って検索をする利用者に向けたシステムといえる。

#### 4.3 利用例

本システムは、現在、電子メール、ニュースなどの検索や、特定テーマに関するインターネット上の検索などに使用されている。

この他に特徴的な利用例として、OCR 読み取りをした文書の検索への適用が挙げられる。各企業は、図表を含む文書を、紙ベース、または、光画像ファイルなどによるイメージの形で情報を保管している。これらの文書をデータベースとして活用したい場合、その内容を再度電子化するコストは大きい。そこで、これらのイメージ情報を OCR 読み取りして、その読み取りデータをそのまま、本システムで索引付けしている。

OCR 読み取りの精度は向上し、精度の高いソフトが市販されている。これらのシステムの地の文の認識率は非常に高く、読み誤りも比較的少ないが、特に、図や表の中やその周辺での認識率は高くないため、OCR 文書には、このような誤認識の結果による誤字・脱字などが避けられない。

このため、OCR 文書は従来の形態素解析技術などが十分機能しない領域であるが、4.1 節〔4〕項で述べた頑健な索引方式と、本システムの曖昧検索機能を制御することで、カバーしている。正確な文書内容は、元のイメージを表示することで読む方式を取っている。

#### 4.4 まとめ

本システムは、利用者とのインタラクションを豊か

にすることで検索結果の質の向上を図っている。本システムにおける曖昧検索など、個々の処理は比較的簡単な処理であり、知的処理の度合いは低い。これは 1 つには、迅速なインタラクションを実現するために、計算量の低いアルゴリズムを使用しなければならないという制約によるが、同時に、制御権を渡された利用者が、自分の制御が結果にどのように影響するかという見通しを得やすくするためでもある。エキスパートシステムにおいては、診断結果の適否を人間が判断するために診断結果の説明機能が重視されたが、利用者に検索の制御権を渡す場合には、情報検索においても同様の説明性や透明性が重要となる。

### 5. 今後の課題

本稿では、ユーザとのインタラクションの面から、現在、商用またはそれに近いレベルで実用化されている文書検索システムの基礎技術を概観した後、2 つの検索システムを紹介した。

両システムとも、検索システムが自動的にを行っている推定処理の限界を、利用者からの適切なフィードバックを得て克服することで、実用的システムとなっている。

しかし、2 つのシステムでは、利用者による制御の直接性が異なる。これは、1 つには、2 つのシステムが対象としている利用者の要求レベルが異なるからだと考えられる。利用者が調べたい情報が比較的明確な場合は、「みつかるぞう」におけるように、直接性の高いインタラクションにより、利用者の意図を極力表現させるのが良いであろうし、ある程度漠然としている場合は、ANATAGONOMY におけるように、利用者からの意識的制御を少なくする方が望ましいであろう。

実際には、この 2 つのモードは独立に存在するのではなく、検索活動の中では共に生じるものである。両方のモードをより自然につなぐインタラクションの実現が重要となろう。

このためには、まず、利用者プロファイルのような利用者モデルの推定技術の精度の向上が重要と考える。特に、時々に変化する利用者の関心を的確に把握する技術が重要であろう [Baladonado 97, Underwood 98]。

また、検索システム自身による検索応答の質を一層向上させる必要がある。このためには、テキストからの情報抽出の精度を向上させることが重要である。情報抽出精度の向上により、利用者の目的に応じたサービスの実現が容易になる [Doorenbos 97, Morohashi 95] と同時に、利用者へ提示するナビゲーション情報として活用することにより、よりの確なインタラクショ

ンを実現できると期待される。

### ◇ 参 考 文 献 ◇

- [Ahlberg 94] Ahlberg, C. and Shneiderman, B.: Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, *Proc. of CHI94*, pp.313-317(1994).
- [Allen 98] Allen, R. B.: Interactive Timelines as Information System Interfaces, *Proc. of International Symposium on Digital Libraries 1995*, pp.175-180(1998).
- [Baladonado 97] Baladonado, Michelle and Winograd, T.: SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests, *Proc. of the Conference on Human Factors in Computing Systems*(1997).
- [Doorenbos 97] Doorenbos, R. B., Etzioni, O. and Weld, D.: A Scalable Comparison-Shopping Agent for the World-Wide Web, *Proc. Agents' 97*(1997).
- [Greene 98] Greene, S., Tanin, E., Plaisant, C. and Shneiderman, B.: The End of Zero-Hit Queries: Query Previews for NASA's Global Change Master Directory, *HCIL Technical report*, CS-TR-3879, UMIACS-TR-98-14(1998).
- [Hearst 95] Hearst, M., Karger, D. R. and Pederson, J.O.: Scatter/Gather as a Tool for the Navigation of Retrieval Results, *Proc. of 1995 AAAI Fall Symp. on Knowledge Navigation* (1995).
- [Hearst 96] Hearst, M. and Pederson, J.O.: Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. *Proc. of ACM/SIGIR 96 Conference* (1996).
- [Hearst 97] Hearst, M. and Karadi, C.: Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy, *Proc. ACM/SIGIR 97 Conference* (1997).
- [井上 97] 井上孝史, 大久保日, 杉崎正之: InfoBee テキスト情報検索技術, *NTT R&D*, Vol.46, No.10 (1997).
- [Lawrence 98] Lawrence, S. and Giles, C. L.: Searching the World Wide Web, *Science Magazine* (1998.4.3).
- [Letsche 97] Letsche, T.A. and Berry, M.W.: Large-Scale Information Retrieval with Latent Semantic Indexing. *Information Sciences - Applications 100*, pp. 105-137 (1997).
- [水口 95] 水口充, 増井俊之, ジョージ ボーデン, 柏木宏一: なめらかなユーザインタフェースによる地図情報検索システム, インタラクティブシステムとソフトウェア III—日本ソフトウェア科学会 WISS'95—, pp.231-240(1995).
- [Morohashi 95] Morohashi, M., Takeda, K., Nomiyama, H. and Maruyama, H.: Information Outlining - Filling the Gap between Visualization and Navigation in Digital Libraries, *Proc. of International Symposium on Digital Libraries 1995*, pp.151-158 (1995).
- [中村 96] 中村篤祥他: プール変数実数多項式による嗜好関数の学習, 第52回情報処理学会全国大会 1-55 (1995).
- [西村 96] 西村英樹他: 重み付き相関ルール導出アルゴリズムによる WWW データ資源の発見, 第7回データ工学ワークショップ (DEWS'96), pp.79-84(1996).
- [Rao 95] Rao, R., et al.: Rich Interaction in the Digital Library, *CACM*, Vol. 38, No.4, pp.29-39 (1995).
- [Rocchio 71] Rocchio, J. J.: Relevance feedback in information retrieval, *The Smart Retrieval System - Experiments in Automatic Document Processing*, pp.313-323 (1971).
- [坂上 96] 坂上秀和, 神場知成, 古関義幸: パーソナル電子新聞 ANATAGONOMY の開発と評価, インタラクティブシステムとソフトウェア IV—日本ソフトウェア科学会 WISS'96—, pp.21-30(1996).
- [Sakagami 97] Sakagami, H. and Kamba, T.: Learning Personal Preferences On Online Newspaper Articles From User Behaviors, *Proc. of Sixth International World Wide Web Conference* (1997).
- [Salton 89] Salton, G.: *Automatic Text Processing*, Addison-Wesley (1989).
- [篠原 96] 篠原靖志: ExtractRequest —利用者への情報開示に基づく検索要求抽出, 情報処理学会ヒューマンインタフェース研究会 96-21 (1996).
- [篠原 98] 篠原靖志: 文書検索システム ExtractRequest における用語分析マップによるフィードバックの評価, 情報処理学会データベース研究会 115-7(1998).
- [Shneiderman 94] Shneiderman, B.: Dynamic queries for visual information seeking, *IEEE Software*, Vol.11, No.6, pp.70-77 (1994).
- [住田 96] 住田, 三池: 知的情報検索の動向, *人工知能学会誌*, Vol.11 No.1(1996).
- [武田 96] 武田英明: ネットワークを利用した知的情報統合, *人工知能学会誌*, Vol.11 No.5(1996).
- [堤 98] 堤富士雄, 篠原靖志: キーワードを2次元平面に配置する文書検索システム, *コンピュータソフトウェア*, Vol.15, No.4 (1998).
- [Underwood 98] Underwood, G.M., Maglio, P.P. and Barrett, R.: User-Centered push for timely information delivery, *Proc. of 7th World Wide Web Conference* (1998).
- [Wove 97] Wove, A. E. and Dreilinger, D.: SAVY-SEARCH —A Metasearch Engine That Learns Which Search Engines to Query, *AI Magazine*, Vol.18, No.2 (1997).

### 著 者 紹 介



篠原 靖志(正会員)

1982年3月東京大学理学部情報科学科卒業。1984年3月東京大学大学院理学研究科情報科学専攻修士課程修了。1984年4月(財)電力中央研究所入所。主として知識獲得, 機械学習, 文書検索に従事。  
<sinohara@criepi.denken.or.jp>



堤 富士雄

1988年3月九州大学工学部情報工学科卒業。1990年3月九州大学大学院工学研究科情報工学専攻修士課程修了。1990年4月(財)電力中央研究所入所。主として演繹データベース, 文書検索, 画像検索研究に従事。  
<tutumi@criepi.denken.or.jp>